(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 922 061 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
23.09.2015 Bulletin 2015/39

(51) Int Cl.:
G11B 27/034 *(2006.01)* ***G11B 27/28*** *(2006.01)*
G06F 17/30 *(2006.01)*

(21) Application number: **15158182.4**

(22) Date of filing: **09.03.2015**

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
MA

(30) Priority: **17.03.2014 JP 2014054207**

(71) Applicant: **FUJITSU LIMITED**
Kawasaki-shi,
Kanagawa 211-8588 (JP)

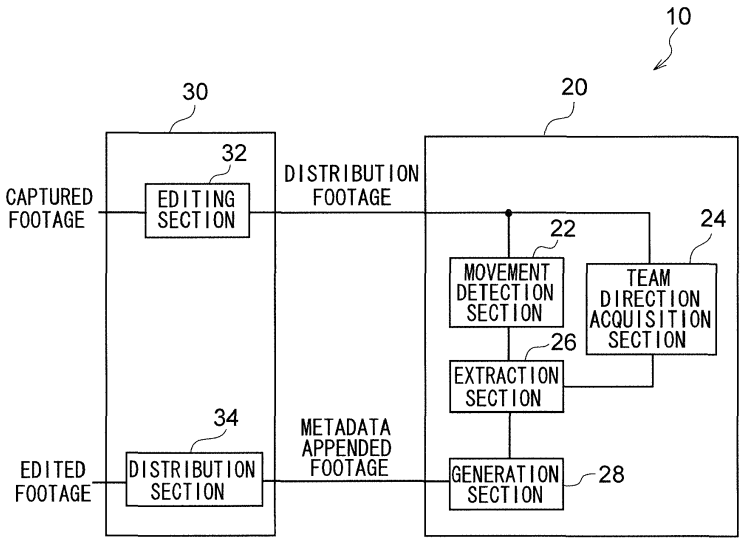
(72) Inventors:
- **Oguchi, Atsushi**
  Kanagawa, 211-8588 (JP)
- **Endo, Susumu**
  Kanagawa, 211-8588 (JP)
- **Baba, Takayuki**
  Kanagawa, 211-8588 (JP)
- **Fujita, Sokichi**
  Kanagawa, 211-8588 (JP)
- **Muramatsu, Tatsuma**
  Kanagawa, 211-8588 (JP)

(74) Representative: **Hoffmann Eitle**
Patent- und Rechtsanwälte PartmbB
Arabellastraße 30
81925 München (DE)

(54) Extraction method and device

(57) An extraction device includes a processor that executes a procedure. The procedure includes: from captured images obtained by capturing a sports game having a match style of attack toward a goal provided in territory of an opposing team, identifying a portion of captured images in which the captured images change over time toward a direction of one or other goal; and extracting a scene in which one or other team is attacking based on the portion of captured images.

FIG.1



EP 2 922 061 A1

Description

FIELD

[0001] The embodiments discussed herein are related to a recording medium storing an extraction program, an extraction method, and an extraction device.

BACKGROUND

[0002] Hitherto, services exist that distribute captured footage of a sports game both as a live distribution, and as a distribution of Video On Demand (VOD). In such services, sometimes there are occasions when the extraction of a particular scene is desired.

[0003] For example, the identification of a scene of interest is performed when a sharp rise in sound volume is detected in audio data corresponding to footage.

[0004] There has also been a proposal to identify highlight segments in video including a frame sequence.

Related Patent Documents

[0005] Japanese National Publication of International Patent Application No. 2008-511186 The extraction of scenes discussed above is performed with the desire, for example, to specify scenes in which one team is attacking. However, scenes in which a sharp rise in sound volume occurs are not always scenes in which one team is attacking.

SUMMARY

[0006] An object of an aspect of the technology disclosed herein is to extract a scene in which one of the teams is attacking, from captured footage of a sports game.

[0007] According to an aspect of the embodiments, a non-transitory recording medium stores an extraction program that causes a computer to execute a process. The process includes: from captured images obtained by capturing a sports game having a match style of attack toward a goal provided in territory of an opposing team, identifying a portion of captured images in which the captured images change over time toward a direction of one or other goal; and extracting a scene in which one or other team is attacking based on the portion of captured images.

BRIEF DESCRIPTION OF DRAWINGS

[0008]

Fig. 1 is a functional block diagram illustrating a schematic configuration of a footage distribution system according to a first exemplary embodiment;
Fig. 2 is a diagram explaining an example of editing;
Fig. 3 is a diagram illustrating an example of a long distance frame;
Fig. 4 is a diagram illustrating an example of a close-up frame;
Fig. 5 is a diagram explaining camera movement;
Fig. 6 is a diagram illustrating an example of an overlay displaying game states in a first half and a second half respectively;
Fig. 7 is a diagram illustrating detected camera movement;
Fig. 8 is a diagram illustrating an example of a metadata file;
Fig. 9 is a schematic block diagram illustrating an example of computer that functions as an extraction device;
Fig. 10 is a schematic block diagram illustrating an example of a computer that functions as a distribution device;
Fig. 11 is a flowchart illustrating an example of editing processing;
Fig. 12 is a flowchart illustrating an example of extraction processing in the first exemplary embodiment;
Fig. 13 is a flowchart illustrating an example of distribution processing;
Fig. 14 is a functional block diagram illustrating a schematic configuration of a footage distribution system according to a second exemplary embodiment;
Fig. 15 is a diagram explaining player movement;
Fig. 16 is a diagram illustrating an example of a goal area image;
Fig. 17 is a diagram illustrating an example of a goal area image;
Fig. 18 is a diagram illustrating an example of an image other than a goal area image;
Fig. 19 is a flowchart illustrating an example of extraction processing in the second exemplary embodiment; and
Fig. 20 is a diagram explaining a continuous period.

DESCRIPTION OF EMBODIMENTS

[0009] Detailed explanation follows below regarding examples of exemplary embodiments of the technology disclosed herein with reference to the drawings. In each of the exemplary embodiments below, explanation is given regarding an example in which an extraction device of the technology disclosed herein is applied to a footage distribution system that distributes captured footage of a soccer match.

First Exemplary Embodiment

[0010] As illustrated in Fig. 1, a footage distribution system 10 according to a first exemplary embodiment includes an extraction device 20 and a distribution device 30. The extraction device 20 and the distribution device 30 are mutually connected through a network. In the footage distribution system 10, footage appended with meta-

data generated by the extraction device 20 is distributed through the network to a specified terminal by the distribution device 30. The extraction device 20 includes a movement detection section 22, a team direction acquisition section 24, an extraction section 26, and a generation section 28. The distribution device 30 includes an editing section 32, and a distribution section 34.

[0011] First, detailed description is given regarding each section of the distribution device 30.

[0012] The editing section 32 acquires captured footage of a soccer match (referred to as "captured footage" hereafter). Note that in the present exemplary embodiment, the footage is media including video data and audio data. The video data included in the captured footage is captured at a frame rate such as 30 fps or 60 fps, and includes plural frames. Each of the frames is associated with time data indicating the elapsed time since capture of the captured footage started. The captured footage includes audio data including audio such as cheering in the match stadium, and audio of commentators, live reports, and the like. The audio data is time sequenced data indicating a level of loudness for respective sampling points, and each sampling point is associated with time data that synchronizes with time data associated to each frame in the video data such that the audio data and the video data are synchronized.

[0013] The editing section 32 applies editing instructions, instructed by an operation by an operator using a display device and an input device, not illustrated in the drawings, to each frame in the video data included in the captured footage using image processing. Editing instructions include, for example, adding an overlay 80 displaying a game state as illustrated in Fig. 2. Below, footage in which editing on the video data included in the captured footage has been performed by the editing section 32 is referred to as "edited footage". The editing section 32 transmits the edited footage to the extraction device 20.

[0014] The distribution section 34 acquires footage to which metadata generated by the extraction device 20 has been appended (referred to as "metadata appended footage" hereafter; detailed description given below). The metadata appended footage is converted to broadcast footage, according to specified standards, and distributed to the distribution destination terminal (omitted from illustration in the drawings) by the distribution section 34.

[0015] Next, detailed description is given regarding each section of the extraction device 20.

[0016] The movement detection section 22 acquires edited footage transmitted from the distribution device 30. The movement detection section 22 detects movement of a camera that captured the video data based on the video data included in the edited footage.

[0017] Specifically, the movement detection section 22 first determines whether each respective frame is a frame captured from a long distance (referred to as a "long distance frame" hereafter), or a frame captured from a short distance (referred to as a "close-up frame" hereafter). This determination may be performed using, for example, an identification model for identifying close-up frames and long distance frames. The identification model may be generated by learning from plural long distance frames as illustrated in Fig. 3, and plural close-up frames as illustrated in Fig. 4.

[0018] For frames determined to be long distance frames, the movement detection section 22 causes parallel translation of data of each pixel in the frame by a specified amount of pixels worth in the horizontal direction as illustrated in Fig. 5, and derives a parallel translation amount for which there is the greatest level of overlap (degree of matching) with a prior frame. For example, taking parallel translation toward the right side of the screen as a positive direction, and parallel translation toward the left side of the screen as a negative direction, the sign of the derived parallel translation amount is the camera movement direction, and the absolute value of the amount of parallel translation is the magnitude of the movement.

[0019] Note that for simplicity of explanation, explanation is given here regarding a case in which movement of the camera in question is limited to the horizontal direction. However, similar detection may be made for vertical direction movement. Moreover, the prior frame for deriving the amount of parallel translation may be the frame one frame prior, or may be a frame a specified number of 2 or more frames prior.

[0020] The movement of the camera that actually captures the captured footage may be measured using hardware, and the measurement amount may be detected as the movement of the camera by acquisition by the movement detection section 22. For example, an encoder may be installed on a platform on which the camera is set up, and the movement of the camera measured using the encoder. In this case the movement detection section 22 may acquire the measurement amount data measured by the encoder together with the captured footage.

[0021] The team direction acquisition section 24 acquires the goal directions in which the respective teams attack for each out of a first half and a second half. For example, when the basic match data is acquired together with the edited footage, the goal directions in which the respective teams attack may be acquired based on this data. Specifically, data such as "at home: Team A vs. away: Team B" may be acquired as the basic data. In this case data such as "in the first half, the home team will attack toward the right side goal direction" may be prepared in advance and matched with the acquired basic data to acquire the goal directions in which the respective teams attack.

[0022] The team direction acquisition section 24 may use data of the overlay added to the edited footage to acquire the goal directions in which the respective teams attack for each out of the first half and the second half. For example, as illustrated in Fig. 6, team names of the respective teams are aligned in the overlay 80 that dis-

plays the game state to correspond to the attaching goal directions of each of the teams for the first half and the second half respectively. For example, when Team B attacks toward the right side goal direction in the first half, as illustrated in Fig. 6, the alignment "Team A - Team B" is displayed in the overlay 80 displaying the game state. However, since Team A becomes the team attacking toward the right side goal direction in the second half, the alignment "Team B - Team A" is displayed in the overlay 80 displaying the game state.

[0023] The team direction acquisition section 24 accordingly selects a frame, out of the video data included in the acquired edited footage, to which the overlay 80 displaying the game state has been added. Then the team direction acquisition section 24 analyzes the content of the overlay 80 displaying the game state in the selected frame using a method such as pattern matching or character recognition, and acquires the goal directions in which each of the respective teams attack for the first half and the second half, respectively.

[0024] Based on the movement of the camera detected by the movement detection section 22, the extraction section 26 determines whether or not the movement is movement toward the goal direction. Movement of the camera in the goal direction can be considered as an attack toward goal by one of the teams. Specifically, the movement of the camera detected by the movement detection section 22 may be represented by a graph with time t as the x axis, and parallel translation amount x as the y axis, as illustrated in Fig. 7. The extraction section 26 is provided with respective threshold values in the positive direction and the negative direction for excluding small camera movements tracking small pass backs ("+th" and "-th" in Fig. 7). Then, the extraction section 26 determines a period in which the parallel translation amount x exceeds the positive threshold value as movement of the camera in the right side goal direction, and determines a period in which the negative threshold value is exceeded as movement of the camera in the left side goal direction.

[0025] The extraction section 26 extracts a group of frames corresponding to a timing in which movement of the camera in the right or left direction has been detected as a scene in which one of the teams attacks (referred to as an "attack scene" hereafter). For example, the extraction section 26 may extract groups of frames associated with time data corresponding to periods in which the parallel translation amount x exceeds the threshold value as groups of frames representing an attack scene, as illustrated in Fig. 7. Groups of frames associated with time data corresponding to periods including a specific number of seconds before and after the period in which the parallel translation amount x exceeds the threshold value may also be extracted as groups of frames that represent an attack scene. This enables attack scenes to be extracted that include a scene leading up to the arrival of the attack, and a scene in which the result of the attack was captured.

[0026] The extraction section 26 determines whether the attack scene represented by the extracted group of frames is a scene from the first half, or a scene from the second half of the game, based on the time data corresponding to the extracted group of frames. Then, the extraction section 26 identifies the attacking team based on the result of the determination as to whether the attack scene is from the first half or the second half, the movement direction of the camera (right or left), and the goal directions in which the respective teams attack for each of the first half and the second half acquired from the team direction acquisition section 24.

[0027] The generation section 28 generates metadata indicating the attack scene based on the group of frames extracted by the extraction section 26. Specifically, the generation section 28 generates metadata associating data indicating the start of the attack scene with time data associated with the leading frame in the extracted group of frames. The generation section 28 also generates metadata associating data indicating the end of the attack scene with time data associated with the final frame of the group of frames extracted by the extraction section 26. Data indicating the attacking team identified by the extraction section 26 is also added to the metadata.

[0028] The generation section 28 generates a metadata file storing plural generated metadata, in the sequence of the time data contained in the metadata. The metadata file may, for example, be generated as a csv (comma-separated values) format file. Fig. 8 illustrates an example of a metadata file. One item of metadata is listed in each row in the example of Fig. 8, and each item of metadata includes time data 82, data indicating the start or end of the attack scene (reference numeral 84), and data indicating the attacking team (reference numeral 86).

[0029] Note that although explanation is given here regarding a case in which the metadata is generated using time data associated with frames, the metadata may be generated using other data that identifies each frame, such as frame numbers.

[0030] The edited footage is appended with the generated metadata file and transmitted to distribution device 30 as metadata appended footage by the generation section 28.

[0031] The extraction device 20 may, for example, be implemented by a computer 40 as illustrated in Fig. 9. The computer 40 includes a CPU 42, memory 44, a non-volatile storage section 46, an input/output interface (I/F) 47, and a network I/F 48. The CPU 42, the memory 44, the storage section 46, the input/output I/F 47, and the network I/F 48 are mutually connected through a bus 49.

[0032] The storage section 46 may be implemented by a hard disk drive (HDD), flash memory, or the like. An extraction program 50 that causes the computer 40 to function as the extraction device 20 is stored in the storage section 46 that serves as a recording medium. The CPU 42 reads the extraction program 50 from the storage section 46 and expands the extraction program 50 into

the memory 44, and sequentially executes the processes included in the extraction program 50.

[0033] The extraction program 50 includes a movement detection process 52, a team direction acquisition process 54, an extraction process 56, and a generation process 58. The CPU 42 operates as the movement detection section 22 illustrated in Fig. 1 by executing the movement detection process 52. The CPU 42 also operates as the team direction acquisition section 24 illustrated in Fig. 1 by executing the team direction acquisition process 54. The CPU 42 also operates as the extraction section 26 illustrated in Fig. 1 by executing the extraction process 56. The CPU 42 also operates as the generation section 28 illustrated in Fig. 1 by executing the generation process 58. The computer 40 executing the extraction program 50 thereby functions as the extraction device 20.

[0034] The distribution device 30 may, for example, be implemented by the computer 60 illustrated in Fig. 10. The computer 60 includes a CPU 62, memory 64, a non-volatile storage section 66, an input/output I/F 67, and a network I/F 68. The CPU 62, the memory 64, the storage section 66, the input/output I/F 67, and the network I/F 68 are mutually connected through a bus 69. A display device and an input device, not illustrated in the drawings, are connected to the computer 60 through the input/output I/F 67.

[0035] The distribution device 30 and the extraction device 20 are connected to the network, through the network I/F 68 of the distribution device 30, and the network I/F 48 of the extraction device 20.

[0036] The storage section 66 may be implemented by a HDD, flash memory, or the like. A distribution program 70 that causes the computer 60 to function as the distribution device 30 is stored in the storage section 66 that serves as a recording medium. The CPU 62 reads the distribution program 70 from the storage section 66 and expands the distribution program 70 into the memory 64, and sequentially executes the processes included in the distribution program 70.

[0037] The distribution program 70 includes an editing process 72 and a distribution process 74. The CPU 62 operates as the editing section 32 illustrated in Fig. 1 by executing the editing process 72. The CPU 62 also operates as the distribution section 34 illustrated in Fig. 1 by executing the distribution process 74. The computer 60 executing the distribution program 70 thereby functions as the distribution device 30.

[0038] The extraction device 20 and the distribution device 30 may each be implemented by, for example, a semiconductor integrated circuit, and more specifically by an application specific integrated circuit (ASIC) or the like.

[0039] Explanation next follows regarding operation of the footage distribution system 10 according to the first exemplary embodiment. When the distribution device 30 is input with captured footage, the distribution device 30 executes editing processing illustrated in Fig. 11, and outputs edited footage. Next, when the extraction device 20 is input with the edited footage, the extraction device 20 executes extraction processing illustrated in Fig. 12, and outputs metadata appended footage. Then, when the distribution device 30 is input with the metadata appended footage, the distribution device 30 executes distribution processing illustrated in Fig. 13, and outputs distribution footage. Each type of processing is described in detail below.

[0040] First, at step S10 of the editing processing illustrated in Fig. 11, the editing section 32 acquires captured footage. Next, at step S 12, the editing section 32 applies editing instructions instructed by operation of an operator using the display device and the input device, not illustrated in the drawings, to each frame of the video data included in the captured footage, using image processing. Next, at step S14, the editing section 32 transmits the edited footage to the extraction device 20, and editing processing ends.

[0041] Next, at step S20 of the extraction processing illustrated in Fig. 12, the movement detection section 22 acquires the edited footage transmitted from the distribution device 30.

[0042] Next, at step S22, the movement detection section 22 determines whether each frame of the video data included in the acquired edited footage is a long distance frame, or a close-up frame. Then, for the frames determined to be long distance frames, the movement detection section 22 parallel-translates data of each pixel in the frame by specified amounts of pixel worth in the horizontal direction, and the parallel translation amount giving the greatest level of overlap with the prior frame is derived as the camera movement direction and magnitude.

[0043] Next, at step S24, based on the basic data of the game acquired together with the captured footage, the team direction acquisition section 24 acquires the goal directions in which the respective teams attack for the first half and the second half, respectively. In cases in which the basic data has not been acquired, the goal directions in which the respective teams attack is acquired for each out of the first half and the second half using data of the overlay 80 displaying the game state added to the edited footage. Note that the present step may be executed at an appropriate timing straight after the edited footage has been input to the extraction device 20, such as at the moment when a frame to which the overlay 80 displaying the game state has been added is first detected. Moreover, the present step may be executed in parallel with the processing of the prior step and the subsequent step.

[0044] Next, at step S26, based on the camera movement detected by the movement detection section 22, the extraction section 26 determines whether or not the movement indicates movement toward a goal direction. Then, the extraction section 26 extracts as an attack scene of one of the teams, a group of frames corresponding to the timing at which the camera movement toward the right or the left direction was detected. The extraction

section 26 further identifies the team that is attacking based on a determination result as to whether the attack scene is from the first half or the second half, the camera movement direction (right or left), and the goal directions in which the respective teams attack for each out of the first half and the second half acquired by the team direction acquisition section 24.

[0045] Next, at step S28, the generation section 28 generates metadata associating data representing the start of the attack scene to time data associated with the leading frame of the group of frames extracted by the extraction section 26. The generation section 28 also generates metadata associating data representing the end of the attack scene to time data associated with the final frame of the group of frames extracted by the extraction section 26. The data indicating the team that is attacking, identified by the extraction section 26, is further added to the metadata.

[0046] Next, at step S30, the generation section 28 generates a metadata file storing the plural metadata generated at step S28 above in the sequence of the time data included in the metadata. The generation section 28 then appends the generated metadata file to the edited footage, and transmits the edited footage to the distribution device 30 as metadata appended footage, and extraction processing ends.

[0047] Next, at step S40 of the distribution processing illustrated in Fig. 13, the distribution section 34 acquires the metadata appended footage transmitted from the extraction device 20. Next, at step S42, the distribution section 34 converts the metadata appended footage to distribution footage according to the specified standards, and distributes the distribution footage to the distribution destination terminal (omitted from illustration in the drawings), and distribution processing ends.

[0048] As explained above, the extraction device 20 according to the first exemplary embodiment extracts as the attack scene, a group of frames corresponding to a timing at which the movement of the camera that captured the captured footage of the soccer game in the goal direction was detected. Extraction of a scene in which one of the teams is attacking is thus enabled.

[0049] Moreover, in cases in which the basic data of the game, or overlay data or the like can be employed to acquire which team is attacking in which goal direction, which team is attacking can also be identified based on the direction (right or left) in which the camera moved.

[0050] Although explanation has been given in the first exemplary embodiment of a case in which the team that is attacking is identified, the processing that is executed by the team direction acquisition section 24 may be omitted, and the attack scene extracted on its own, without identifying the team that is attacking.

Second Exemplary Embodiment

[0051] Explanation next follows regarding a second exemplary embodiment. Note that parts that are similar to those of the footage distribution system 10 according to the first exemplary embodiment are appended with the same reference numerals, and detailed explanation thereof is omitted.

[0052] As illustrated in Fig. 14, a footage distribution system 210 according to the second exemplary embodiment includes an extraction device 220 and a distribution device 30. The extraction device 220 includes a movement detection section 222, a goal area image detection section 25, an extraction section 226, and a generation section 28.

[0053] The movement detection section 222, similarly to the movement detection section 22 of the first exemplary embodiment, determines long distance frames, and detects camera movement based on the long distance frames. The movement detection section 222 then excludes the effect of the detected camera movement, and detects player movement from inter-frame difference images. Excluding the effect of the camera movement refers to excluding static background portions appearing as moving bodies in the difference images due to the movement of the camera when the inter-frame difference images are generated.

[0054] Specifically, the movement detection section 222 generates an inter-frame difference image between two frames in which data of each of the pixels in one of the frames have been parallel-translated by the amount of camera movement between the frames. Moving bodies such as the players and the ball thereby appear as small regions in the difference image. The movement direction and the movement magnitude of each of the moving bodies can be detected by making associations between each body in the difference images as illustrated in Fig. 15 for example.

[0055] Note that when generating a difference image between two frames, the difference images may be generated between consecutive frames, or, for example, a difference image may be generated between frames a specified amount of frames apart, such as every 0.5 seconds to 1 second. Since the magnitude of the camera movement between frames becomes large in the latter case, this enables a reduction in error compared to cases in which small camera movements are detected.

[0056] The goal area image detection section 25 acquires the edited footage transmitted from the distribution device 30, and detects frames representing images capturing a goal area (referred to as "goal area images" hereafter) from each frame of the video data included in the edited footage. Specifically, the goal area image detection section 25 extracts an image characteristic from each frame. The image characteristic is a characteristic represented by pixel data, such as brightness and color, of each pixel in the frame. Then, the goal area image detection section 25 derives degrees of similarity between a reference image characteristic, representing a pre-prepared goal area image, and the extracted image characteristics extracted from each frame of the edited footage. The goal area image detection section 25 detects as

frames representing goal area images frames with an extracted image characteristic having a degree of similarity to the reference image characteristic that is a specified value or above.

[0057] More specifically, the goal area image is, for example, an image such as that illustrated in Fig. 16, and has a characteristic such as a goal post being present in a portion of the image and there being a group of pixels of a color different from the color of the pitch in a corner portion of the image (inside the dashed line in Fig. 16). Accordingly, the goal area image detection section 25, for example, excludes pixels representing the color of the pitch (green) based on color data of the pixels in each of the frames. Moreover, pixels are grouped based on color data, and small regions in which the number of pixels included in the group is a specific number or fewer, are excluded as regions representing players. Then, the frame is detected as a goal area image in cases in which a characteristic represented by a position and shape of the remaining regions and colors of the pixels included in the remaining regions indicate, for example, a characteristic of a group of pixels of a color different from the color of the pitch in a corner portion of the image, as illustrated inside the dashed line in Fig. 16.

[0058] The goal area image detection section 25 may determine whether or not each frame is a goal area image by pattern matching of the shape of a goal post, a line in front of the goal post, or the like. Although Fig. 16 illustrates a goal area image representing the goal area at the right side of the screen, goal area images representing the goal area at the left side of the screen, as illustrated in Fig. 17, are detected similarly.

[0059] As the image characteristic, the goal area image detection section 25 may employ a characteristic representing an image structure (composition), a characteristic representing a placement or distribution of, for example, color and brightness, a characteristic obtained by wavelet conversion, or the like. In cases in which such an image characteristic is extracted, it may be extracted from the entirety of each frame, or the characteristic may be extracted from a portion of each frame.

[0060] Based on the player movement detected by the movement detection section 222, the extraction section 226 determines whether or not that movement represents movement in a goal direction. A player moving in a goal direction can be considered as an indication that one of the teams is attacking toward a goal. Specifically, as illustrated in Fig. 7, the player movement detected by the movement detection section 222 can be represented by a graph the time t as the x axis, and the horizontal direction translation amount x as the y axis, similarly to the camera movement in the first exemplary embodiment. In the present exemplary embodiment, attention is given to the amount of translation in the horizontal direction in the detected player movements, since player movement in a goal direction is important data. Note that the translation amount in the vertical direction may be used in combination. Since horizontal direction translation amounts are obtained for each of the plural players, a single horizontal direction translation amount is derived for each team using an average of all of the obtained horizontal direction translation amounts, or the like.

[0061] The extraction section 26 is provided with a threshold value for each out of the positive direction and the negative direction ("+th" and "-th" in Fig. 7) in order to exclude small player movements due to small pass backs. Then, the extraction section 226 determines periods in which the horizontal direction translation amount x exceeds the positive threshold value as player movement in the right side goal direction, and determines periods in which the negative threshold value is exceeded as player movement in the left side goal direction. The extraction section 226 then extracts as an attack scene candidate a group of frames corresponding to a timing at which player movement in the right or left direction was detected.

[0062] Here, as illustrated in Fig. 16 and Fig. 17, images in attack scenes are generally goal area images. For example, as illustrated in Fig. 18, times when images are of the central area of the pitch, can be considered to be scenes in which neither of the teams are attacking in a goal direction. Namely, even when player movement in a goal direction is detected, determination that that scene is not an attack scene is enabled in cases in which a goal area image is not included in the group of frames corresponding to that timing.

[0063] Thus the extraction section 226 determines whether or not frames detected as goal area images by the goal area image detection section 25 are included in at least a portion of the group of frames extracted as an attack scene candidate. The extraction section 226 extracts the attack scene candidate as an attack scene in cases in which a goal area image is included therein. The extraction section 226 may also extract as an attack scene a group of frames in which the extracted goal direction in the goal area image matches the player movement direction out of the groups of frames with goal area images extracted as attack scene candidates. Specifically, for example, a group of frames including a goal area image representing the goal area on the right side of the screen as the goal area image, and in which player movement in the right side goal direction is detected, may be extracted as an attack scene. A group of frames including a goal area image representing the goal area on the left side of the screen as the goal area image, and in which player movement in the right side goal direction is detected, may be excluded from the attack scenes. In cases in which the extracted goal direction of the goal area image and the player movement direction do not match, the attack scene candidate can, for example, be considered to be a scene of a goal kick, or a scene in which a defender clears the ball. Attack scene candidates in which the extracted goal direction of the goal area image and the player movement direction do not match are excluded from the attack scenes. This enables the detection as an attack scene of scenes of goal kicks, and

scenes in which a defender clears the ball, to be suppressed.

[0064] The extraction device 220 may be implemented by, for example, the computer 40 illustrated in Fig. 9, similarly to in the first exemplary embodiment. An extraction program 250 that causes the computer 40 to function as the extraction device 220 is stored in the storage section 46 serving as a recording medium. The CPU 42 reads the extraction program 250 from the storage section 46, expands the extraction program 250 into the memory 44, and sequentially executes the processes included in the extraction program 250.

[0065] The extraction program 250 includes a movement detection process 252, a goal area image detection process 55, an extraction process 256, and the generation process 58. The CPU 42 operates as the movement detection section 222 illustrated in Fig. 14 by executing the movement detection process 252. The CPU 42 also operates as the goal area image detection section 25 illustrated in Fig. 14 by executing the goal area image detection process 55. The CPU 42 also operates as the extraction section 226 illustrated in Fig. 14 by executing the extraction process 256. The generation process 58 is similar to that of the first exemplary embodiment. The computer 40 executing the extraction program 250 thereby functions as the extraction device 220.

[0066] Explanation next follows regarding the operation of the footage distribution system 210 according to the second exemplary embodiment. In the second exemplary embodiment, since only the extraction processing executed by the extraction device 220 is different from the first exemplary embodiment, explanation follows regarding the extraction processing in the second exemplary embodiment with reference to Fig. 19. Note that processing similar to the extraction processing of the first exemplary embodiment is appended with the same reference numerals and detailed explanation thereof is omitted.

[0067] At step S23 of the extraction processing illustrated in Fig. 19, the movement detection section 222 determines long distance frames based on the video data included in the edited footage, and detects camera movement based on the long distance frames. Then, the movement detection section 222 excludes the effect of the detected camera movement and detects player movement from inter-frame difference images.

[0068] Next, at step S25, the goal area image detection section 25 extracts an image characteristic for each frame of the video data, and extracts frames representing a goal area image by comparison with a reference image characteristic representing a pre-prepared goal area image.

[0069] Next, at step S27, the extraction section 226 extracts as an attack scene candidate a group of frames corresponding to a timing at which player movement was detected in the right or the left direction by the movement detection section 222. Then, the extraction section 226 determines whether or not frames detected as goal area images by the goal area image detection section 25 are included in at least a portion of the group of frames extracted as the attack scene candidate. In cases in which a goal area image is included, the extraction section 226 extracts that attack scene candidate as an attack scene.

[0070] Subsequent processing is performed similarly to the extraction processing of the first exemplary embodiment, and the extraction processing ends.

[0071] As explained above, according to the extraction device 220 according to the second exemplary embodiment, a group of frames corresponding to a timing at which player movement in a goal direction was detected is extracted as an attack scene from captured footage of a soccer game. Extraction of a scene in which one of the teams is attacking is thus enabled.

[0072] Moreover, also employing a goal area image detection result in the attack scene extraction enables extraction of scenes that are not attack scenes, based on movement of players in midfield defending or the like, to be prevented.

[0073] Note that although explanation has been given of a case in which player movement is detected in the second exemplary embodiment, ball movement may be detected using a similar method. Detection of ball movement is enabled based on shape and color data of moving bodies detected using difference images.

[0074] Moreover, the team direction acquisition section 24 of the first exemplary embodiment may be provided, and identification of the attacking team may be performed in the second exemplary embodiment also.

[0075] Moreover, the goal area image detection section 25 of the second exemplary embodiment may be provided, and a goal area image detection result may also be employed in the attack scene extraction in the first exemplary embodiment also.

[0076] In each of the exemplary embodiments above, easy location of a desired attack scene of sports game footage (the captured footage, or the edited footage) based on metadata is enabled when employing metadata appended footage, to which the metadata indicating the attack scene extracted is appended. Moreover during footage distribution, for example, a supplementary service, such as transmission of email to a user, may be performed automatically coordinated with attack scenes based on the metadata.

[0077] Note that although explanation has been given regarding cases in which attack scenes are extracted based on video data in each of the above exemplary embodiments, audio data corresponding to the video data may also be employed in order to identify attack scenes. For example, scenes of interest may be extracted based on the audio data, and when camera or player movement in a goal direction is detected that is also a scene of interest, that scene may be extracted as an attack scene.

[0078] Specifically, a sound change detection section may be provided, and changes calculated in the sound volume level of the audio data included in the edited footage. For example, the difference or ratio of an average

value of the sound volume level at respective sampling points included in a specific interval, and the average value of the sound volume level at respective sampling points included in a specific interval immediately before, may be calculated as the sound volume change amount. The sound change detection section then detects as start points, points in time (sampling points) at which the calculated amount of change in the sound volume exceeds a predetermined change amount threshold value th_c.

[0079] For example, the volume of cheering and the like usually rises suddenly in the game stadium at the start time of scenes of interest such as goal scenes, foul scenes, scenes in which defense switches to attack such as at a successful intercept or the like, and disappointed near miss scenes. Accordingly, an interval short enough that sudden changes to the sound volume level can be captured is set as the specific interval when calculating sound volume level changes. The change amount threshold value the is preset based on the amount of sound volume level change in audio data corresponding to scenes of interest in past game footage.

[0080] The extraction section 26 (or 226) determines a continuous period in which the sound volume is continually a specific level or greater from a start point detected by the sound change detection section. The volume of the cheering or the like tends to be continuous for a specific period or longer after suddenly rising in scenes of interest. Accordingly, the continuous interval, at which the sound volume is the specific level or greater, is determined such that such a trend can be captured. The specific level for identification of the continuous period may, as illustrated in Fig. 20 for example, be a level M1 that is a specific percentage (for example 50%) of a level M0 of the sound volume at a start point t0. In this case, the extraction section 26 identifies a period from the start point t0 until a point t1 at which the level of the sound volume became less than M1 as the continuous period (continuous period A in Fig. 20). As another example, as illustrated in Fig. 20 for example, the specific level may be a level M2 of the sound volume immediately before the start point t0, namely, before the sudden rise in the sound volume. In this case, the extraction section 26 identifies as the continuous period a period from the start point t0 until a point t2 at which the level of the sound volume becomes less than M2 (continuous period B in Fig. 20).

[0081] The extraction section 26 (or 226) compares the identified continuous period against a preset continuous period threshold value th_t. In cases in which the continuous period is the continuous period threshold value th_t or greater, the extraction section 26 (or the 226) extracts a frame of the video data corresponding to the start point detected by the sound change detection section as a starting frame of a scene of interest. The extraction section 26 (or 226) extracts as an attack scene candidate a group of frames corresponding to a timing at which movement of the camera or the players in a goal direction was detected. The extraction section 26 (or 226) then extracts that attack scene candidate as an attack scene when a least a portion of the frames of the attack scene candidate are also extracted as a scene of interest. This thereby enables extraction of scenes of back passes and the like as attack scenes to be prevented.

[0082] Although explanation has been given in which soccer is the target sport in each of the exemplary embodiments above, there is no limitation thereto. For example, technology disclosed herein may be applied to any sport, such as rugby or American football, having a match structure with attack towards a goal provided in territory of an opposing team.

[0083] Although explanation has been given of examples in which a footage distribution system includes a distribution device and an extraction device in each of the exemplary embodiments above, there is no limitation thereto. Each of the functional sections of the distribution device, and each of the functional sections of the extraction device may be implemented by a single computer.

[0084] Although explanation has been given of cases in which footage appended with metadata, generated by the extraction device, indicating attack scenes is distributed by the distribution device in each of the exemplary embodiments above, there is no limitation thereto. For example, metadata appended footage may be saved as a large volume archive, and the desired attack scenes alone extracted and output based on the metadata.

[0085] The output from the extraction device may also be employed in applications other than footage distribution systems.

[0086] Although explanation has been given above of modes in which the extraction programs 50 and 250 are pre-stored (installed) on the storage section 46, and the distribution program 70 is pre-stored (installed) on the non-volatile storage section 66, they may be provided in a format recorded on a recording medium such as a CD-ROM or a DVD-ROM.

[0087] One aspect exhibits the advantageous effect of enabling extraction of scenes of one of the teams attacking from captured footage of a sports game.

Claims

1. An extraction program that causes a computer (40) to execute processing comprising:

   from captured images obtained by capturing a sports game having a match style of attack toward a goal provided in territory of an opposing team, identifying (52) a portion of captured images in which the captured images change over time toward a direction of one or other goal; and extracting (56) a scene in which one or other team is attacking based on the portion of the captured images.

2. The extraction program of claim 1, the processing

further comprising:

detecting (55) a frame capturing a goal area from each of a plurality of frames included in the captured images, wherein
when extracting the scene in which one or other of the teams is attacking, in cases in which the frame capturing the goal area is included in a group of frames corresponding to a timing at which movement toward the direction of one or other of the goals was detected, the group of frames is extracted as the scene in which one or other of the teams is attacking.

3. The extraction program of claim 2, wherein, in the processing, when extracting the scene in which one or other of the teams is attacking, a group of frames in which a direction of movement toward the direction of one or other of the goals, and a detection direction of the goal area in the frame capturing the goal area, match each other is extracted as the scene in which one or other of the teams is attacking.

4. The extraction program of claim 2 or claim 3, wherein, in the processing, the extraction of the scene in which one or other of the teams is attacking is performed by detecting, as the frame capturing the goal area, a frame in which an image characteristic extracted from the frame resembles a characteristic predetermined as a characteristic of a frame capturing the goal area.

5. The extraction program of any one of claim 1 to claim 4, the processing further comprising:

acquiring (54) data of a goal direction in which each team attacks; and
during extraction of the scene in which one or other of the teams is attacking, extracting the attack scene and identifying the team attacking in the attack scene based on detected movement toward the direction of one or other of the goals and the acquired goal direction in which each team attacks.

6. The extraction program of any one of claim 1 to claim 5, wherein, in the processing, a leading frame of a group of frames corresponding to a timing at which movement toward the direction of one or other of the goals was detected is set as a frame associated with time data of a specific number of seconds before the timing.

7. A extraction method comprising:

based on historical data of movement during image capture by a camera that has captured a sports game having a match style of attack toward a goal provided in territory of an opposing team, identifying a portion of captured images representing a change in movement of the camera toward a direction of one or other goal; and
by a processor (42), extracting a scene in which one or other team is attacking based on the portion of the captured images.

8. The extraction method of claim 7, further comprising:

detecting a frame capturing a goal area from each of a plurality of frames included in the captured images, wherein
when extracting the scene in which one or other of the teams is attacking, in cases in which the frame capturing the goal area is included in a group of frames corresponding to a timing at which movement toward the direction of one or other of the goals was detected, the group of frames is extracted as the scene in which one or other of the teams is attacking.

9. The extraction method of claim 8, wherein, when extracting the scene in which one or other of the teams is attacking, a group of frames in which a direction of movement toward the direction of one or other of the goals, and a detection direction of the goal area in the frame capturing the goal area, match each other is extracted as the scene in which one or other of the teams is attacking.

10. The extraction method of claim 8 or claim 9, wherein the extraction of the scene in which one or other of the teams is attacking is performed by detecting, as the frame capturing the goal area, a frame in which an image characteristic extracted from the frame resembles a characteristic predetermined as a characteristic of a frame capturing the goal area.

11. The extraction method of any one of claim 7 to claim 10, the extraction process further comprising:

acquiring data of a goal direction in which each team attacks; and
extracting the attack scene and identifying the team attacking in the attack scene based on detected movement toward the direction of one or other of the goals and the acquired goal direction in which each team attacks.

12. The extraction method of any one of claim 7 to claim 11, wherein a leading frame of a group of frames corresponding to a timing at which movement toward the direction of one or other of the goals was detected is set as a frame associated with time data of a specific number of seconds before the timing.

13. An extraction device, comprising:

a detection section (22) that, from captured images obtained by capturing a sports game having a match style of attack toward a goal provided in territory of an opposing team, detects a presence of an imaging subject moving toward a direction of one or other goal; and
an extraction section (26) that extracts a scene in which one or other team is attacking according to the presence of the detected imaging subject.

14. The extraction device of claim 13, wherein the imaging subject is one or a plurality of players in the sports game, or a ball used in the sports game, or both.

15. The extraction device of claim 14, wherein the detection section excludes an effect of movement of a camera that has captured the game, and detects the imaging subject.

FIG.1

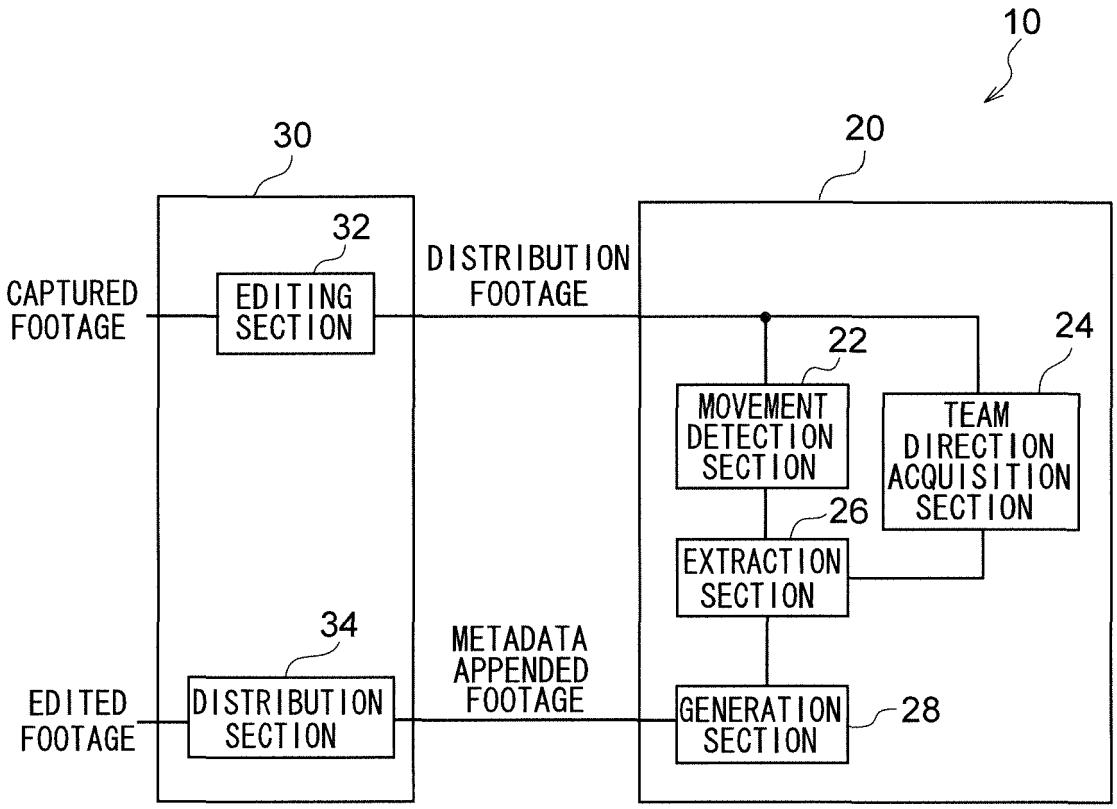

10
30
20
32
CAPTURED FOOTAGE
EDITING SECTION
DISTRIBUTION FOOTAGE
22
24
MOVEMENT DETECTION SECTION
TEAM DIRECTION ACQUISITION SECTION
26
EXTRACTION SECTION
34
EDITED FOOTAGE
DISTRIBUTION SECTION
METADATA APPENDED FOOTAGE
GENERATION SECTION
28

FIG.2

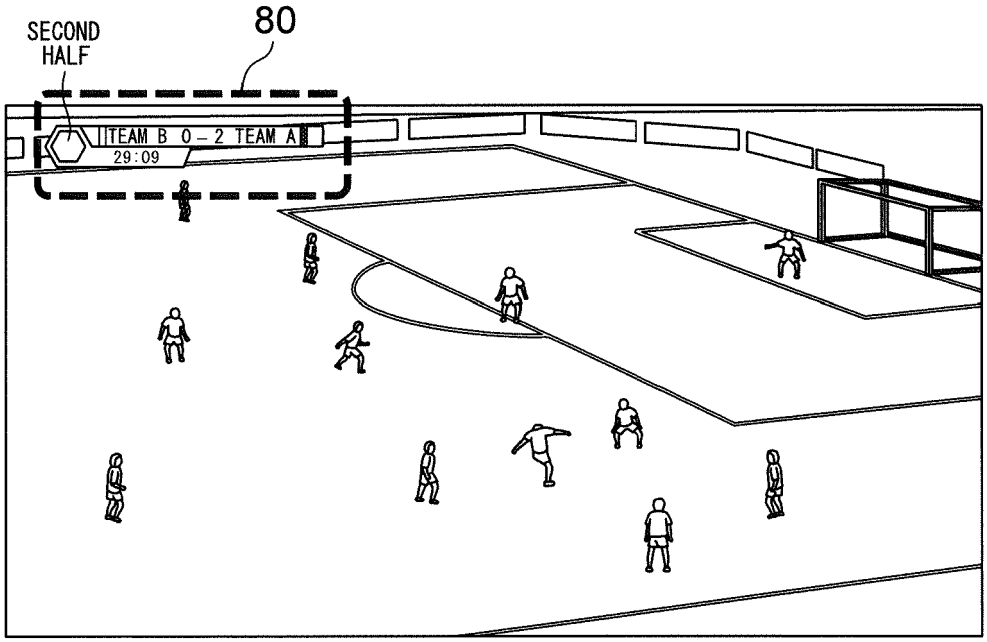


FIG.3

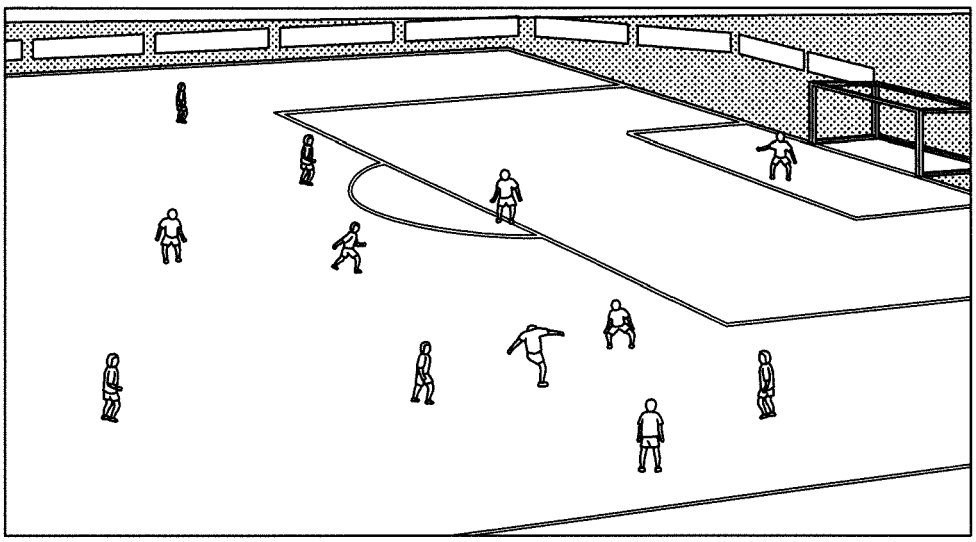


FIG.4

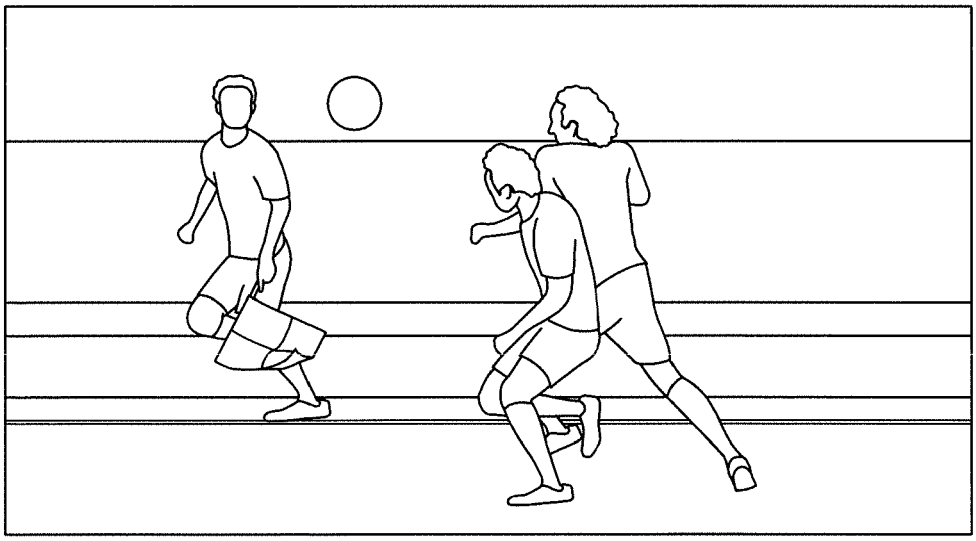


FIG.5

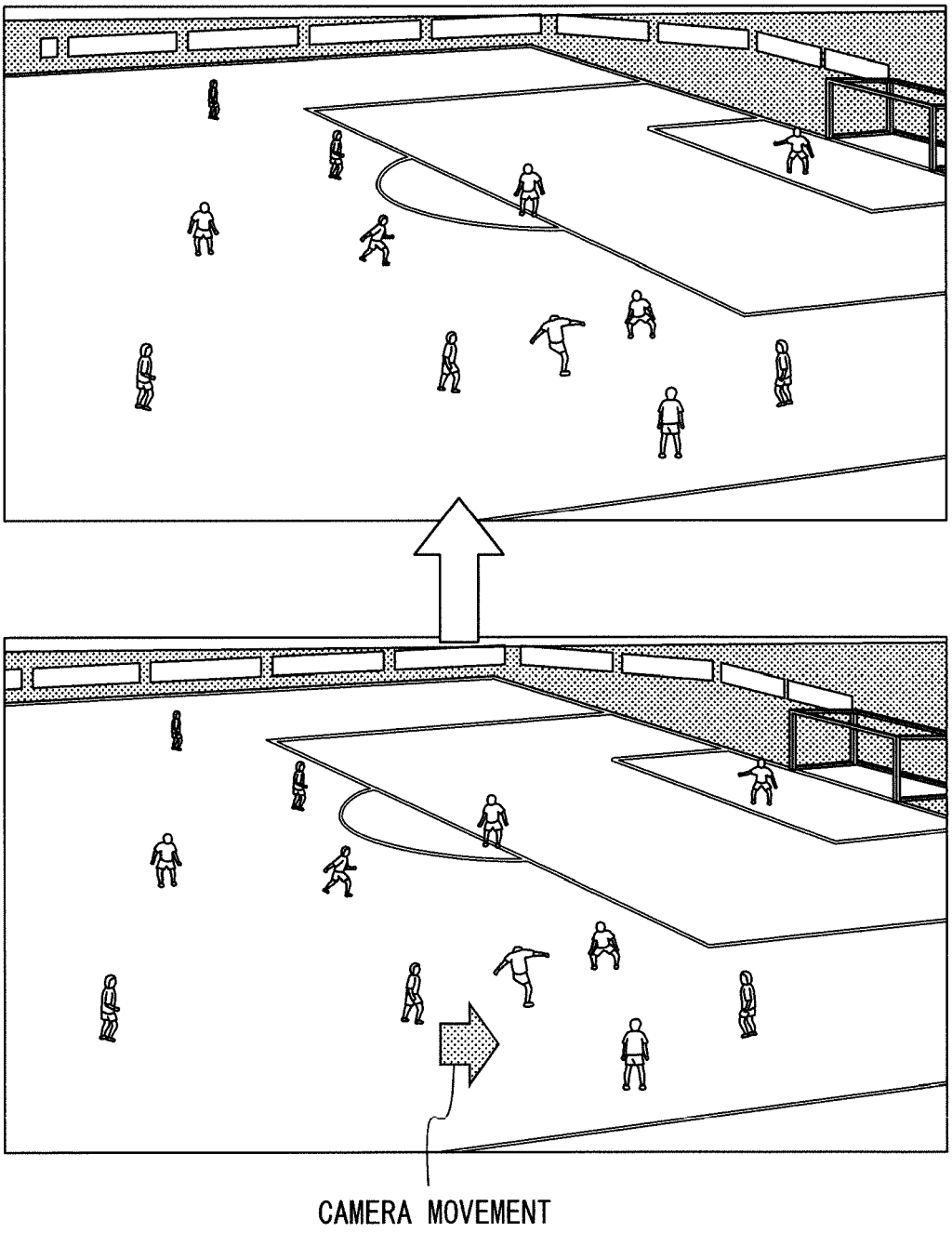


FIG.6

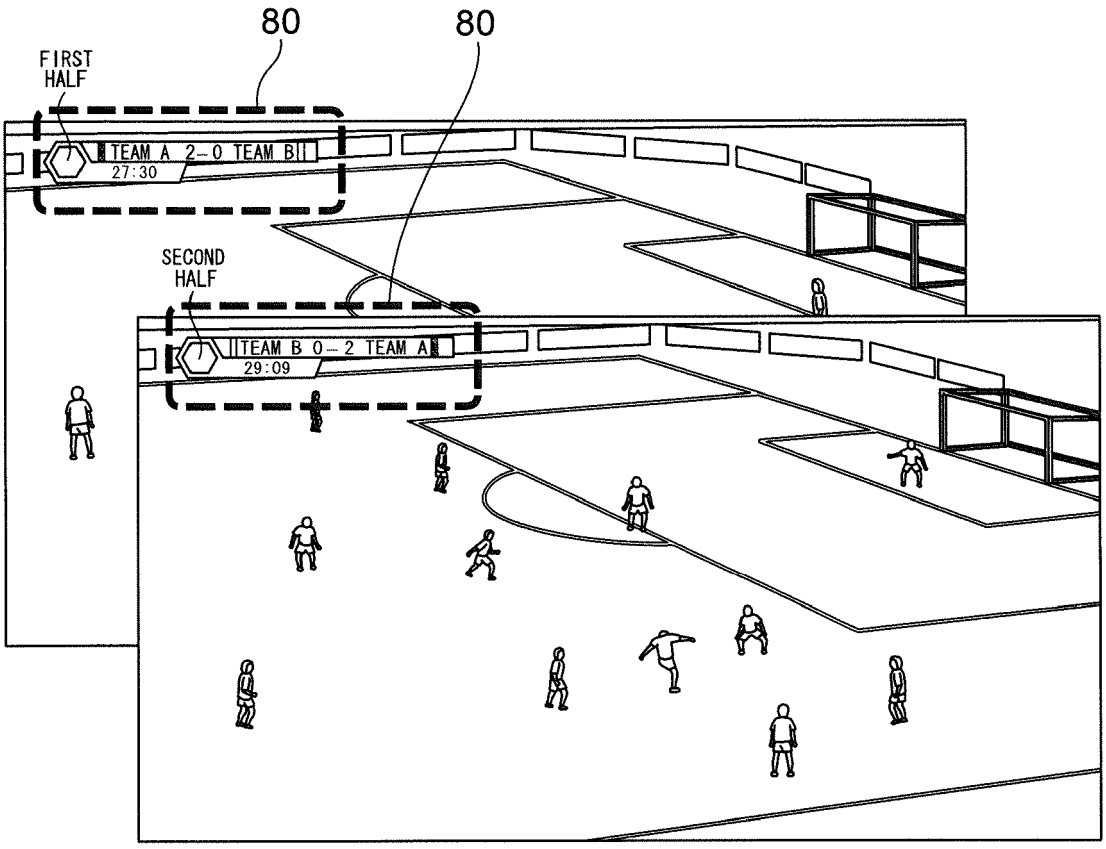

80
80
FIRST HALF
TEAM A 2 – 0 TEAM B
27:30
SECOND HALF
TEAM B 0 – 2 TEAM A
29:09

FIG.7

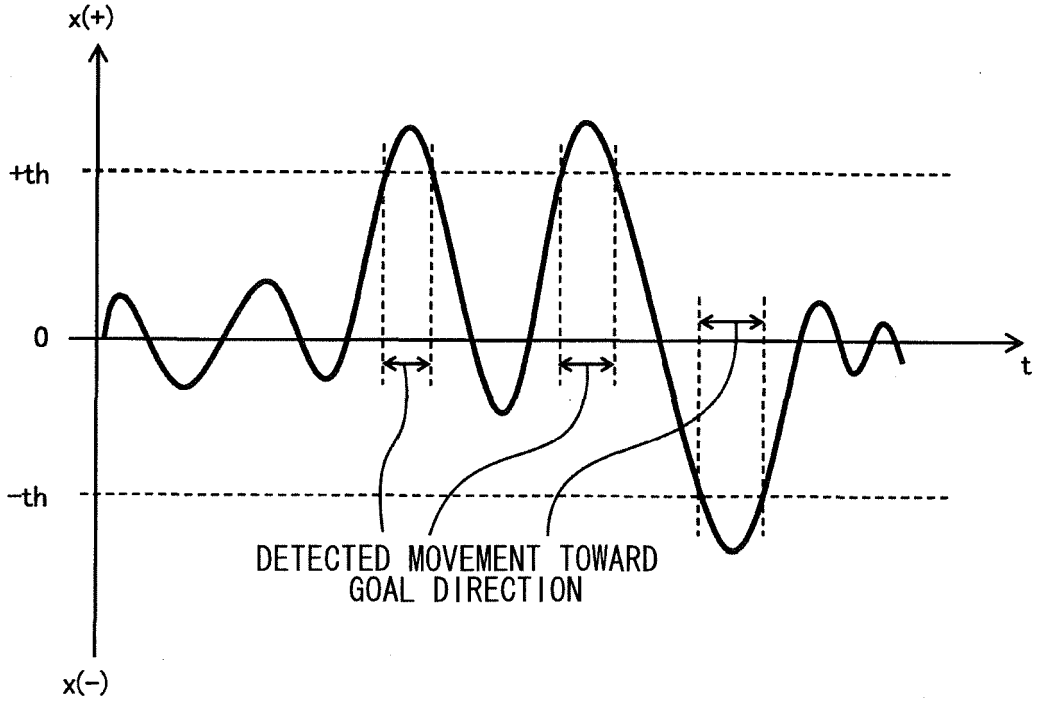


FIG.8

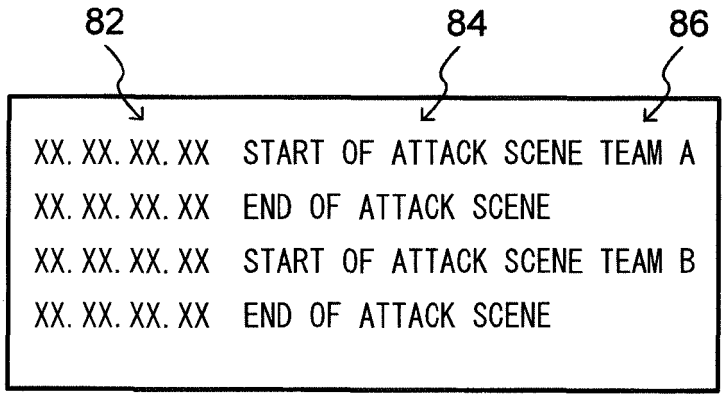

82
84
86
XX.XX.XX.XX START OF ATTACK SCENE TEAM A
XX.XX.XX.XX END OF ATTACK SCENE
XX.XX.XX.XX START OF ATTACK SCENE TEAM B
XX.XX.XX.XX END OF ATTACK SCENE

FIG.9

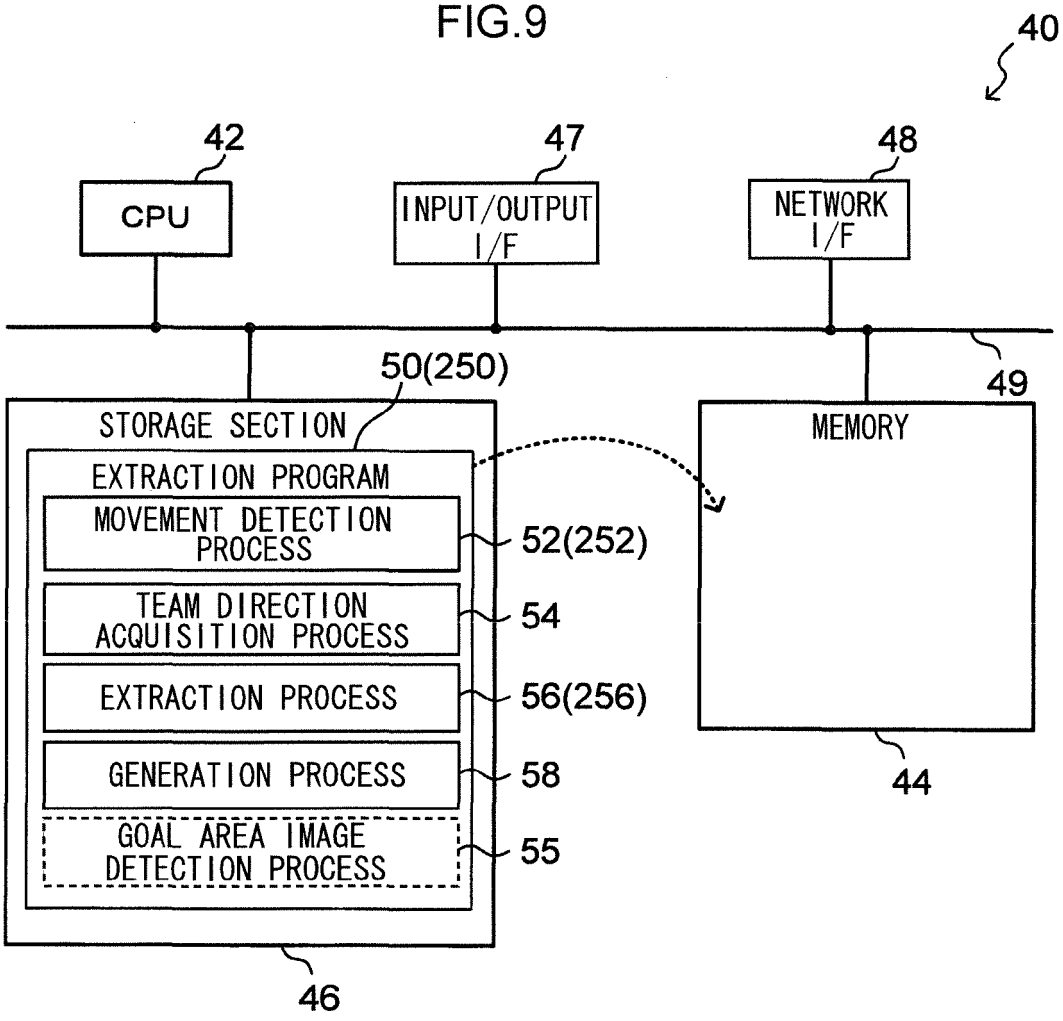

40
42
CPU
47
INPUT/OUTPUT I/F
48
NETWORK I/F
49
50(250)
STORAGE SECTION
EXTRACTION PROGRAM
MOVEMENT DETECTION PROCESS
52(252)
TEAM DIRECTION ACQUISITION PROCESS
54
EXTRACTION PROCESS
56(256)
GENERATION PROCESS
58
GOAL AREA IMAGE DETECTION PROCESS
55
46
MEMORY
44

FIG.10

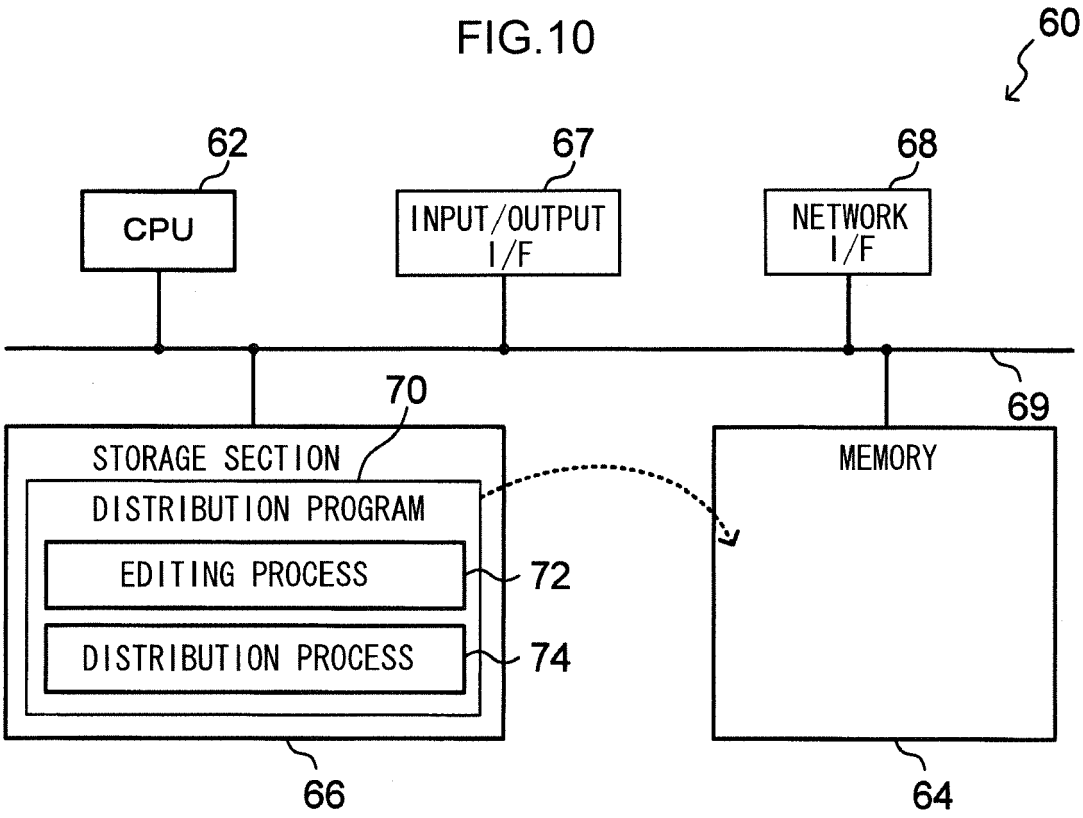

60
62
CPU
67
INPUT/OUTPUT I/F
68
NETWORK I/F
69
70
STORAGE SECTION
DISTRIBUTION PROGRAM
EDITING PROCESS
72
DISTRIBUTION PROCESS
74
MEMORY
66
64

FIG.11

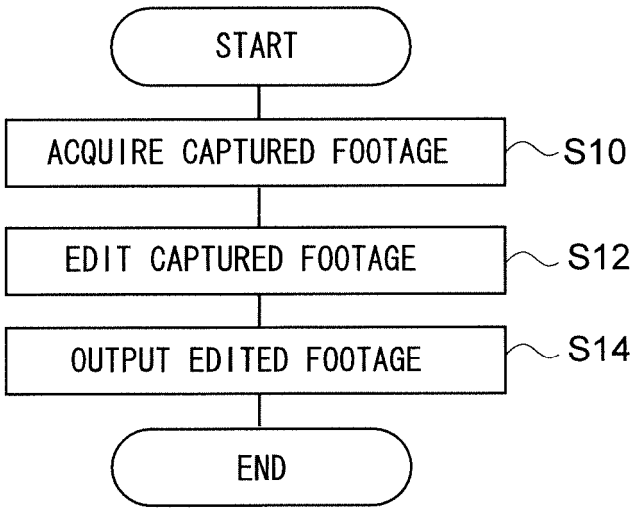

START
ACQUIRE CAPTURED FOOTAGE
S10
EDIT CAPTURED FOOTAGE
S12
OUTPUT EDITED FOOTAGE
S14
END

FIG.12

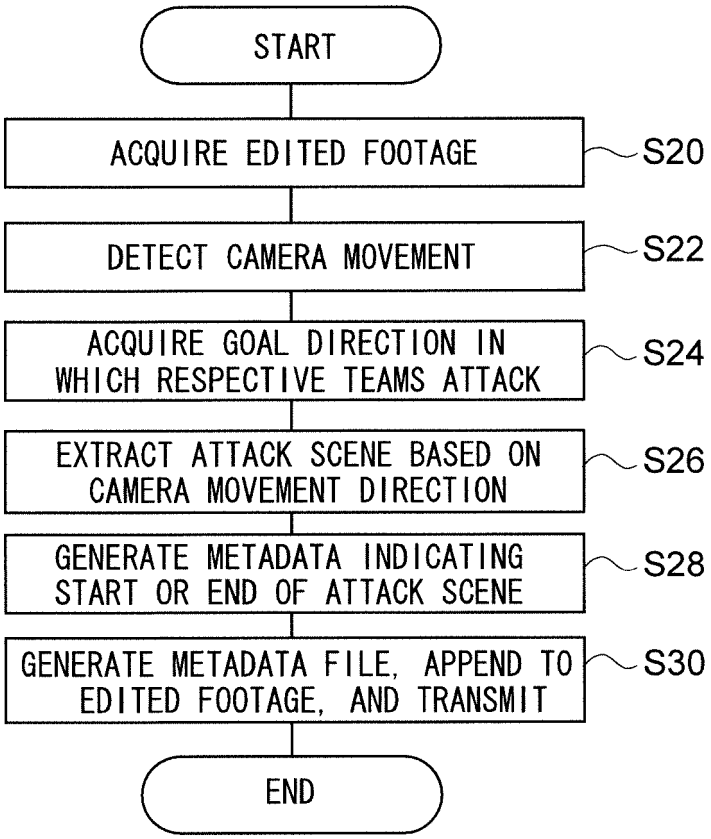

START
ACQUIRE EDITED FOOTAGE
S20
DETECT CAMERA MOVEMENT
S22
ACQUIRE GOAL DIRECTION IN WHICH RESPECTIVE TEAMS ATTACK
S24
EXTRACT ATTACK SCENE BASED ON CAMERA MOVEMENT DIRECTION
S26
GENERATE METADATA INDICATING START OR END OF ATTACK SCENE
S28
GENERATE METADATA FILE, APPEND TO EDITED FOOTAGE, AND TRANSMIT
S30
END

FIG.13

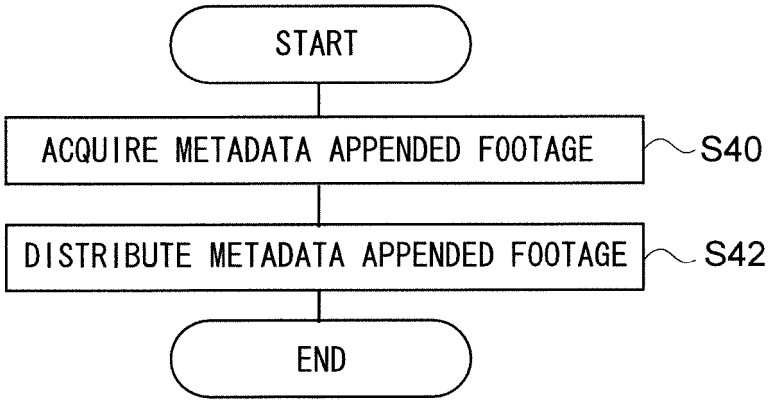

START
ACQUIRE METADATA APPENDED FOOTAGE
S40
DISTRIBUTE METADATA APPENDED FOOTAGE
S42
END

FIG.14

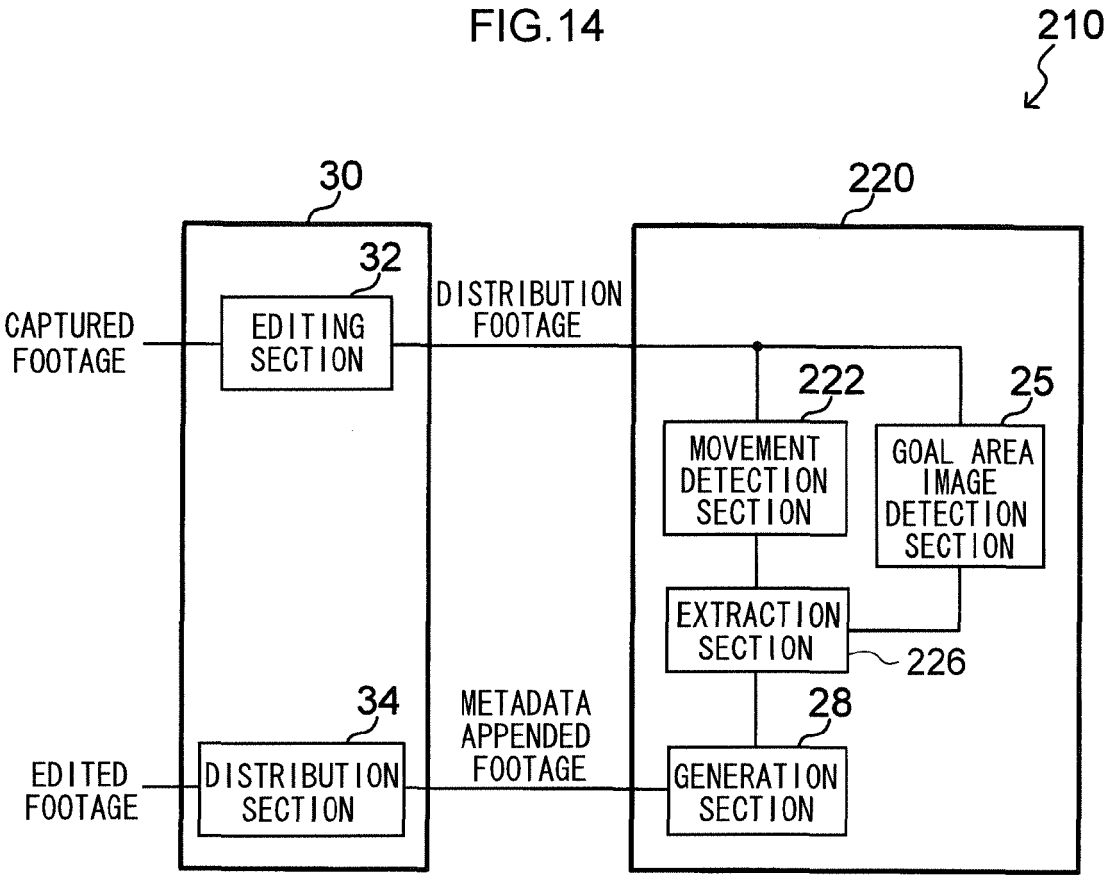

210
30
220
32
CAPTURED FOOTAGE
EDITING SECTION
DISTRIBUTION FOOTAGE
222
25
MOVEMENT DETECTION SECTION
GOAL AREA IMAGE DETECTION SECTION
EXTRACTION SECTION
226
34
METADATA APPENDED FOOTAGE
28
EDITED FOOTAGE
DISTRIBUTION SECTION
GENERATION SECTION

FIG.15

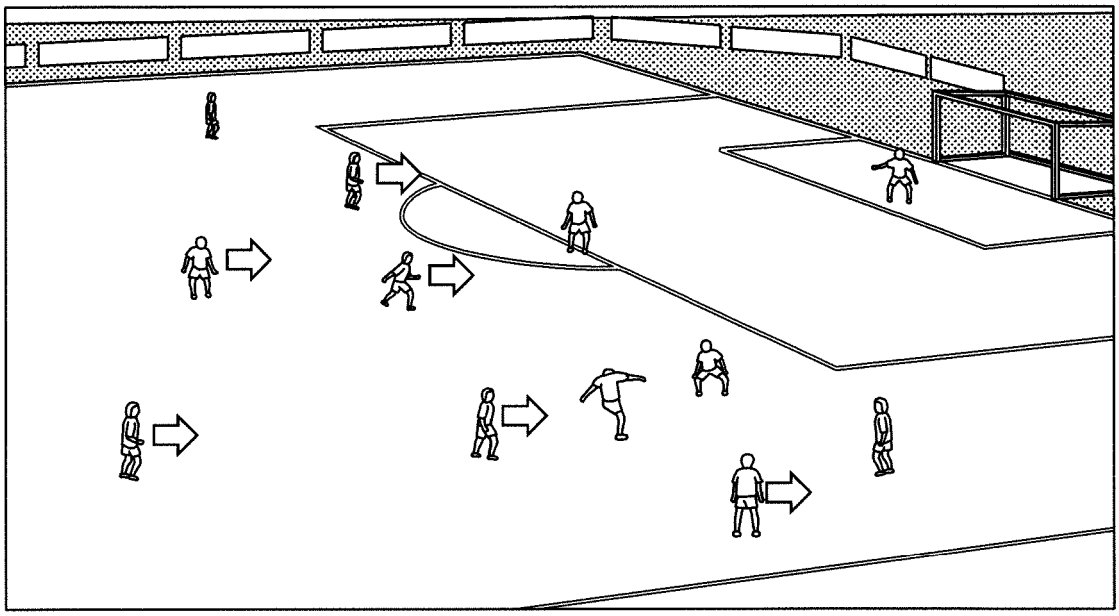


FIG.16

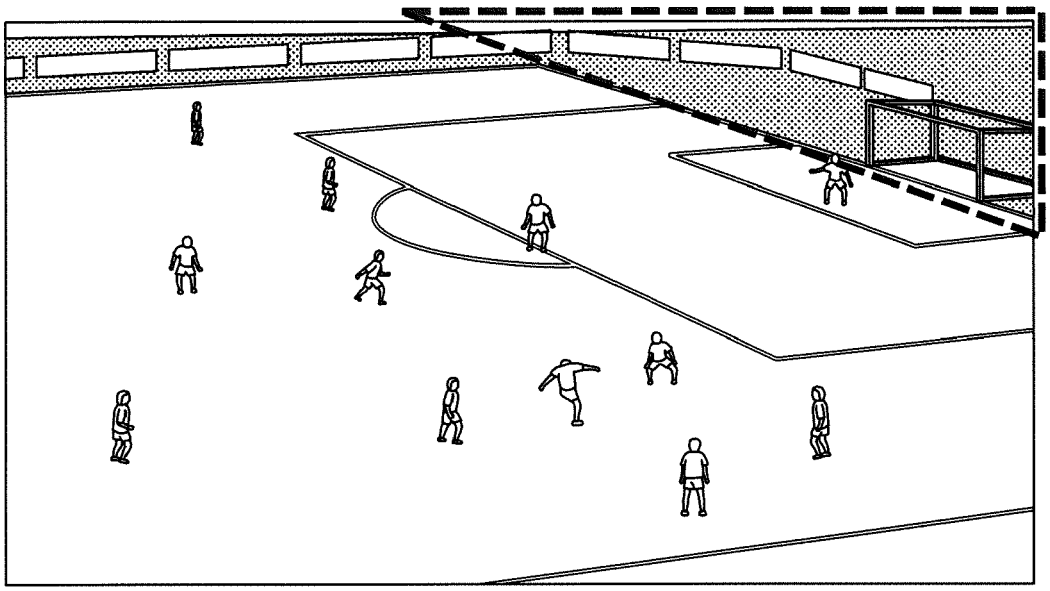


FIG.17

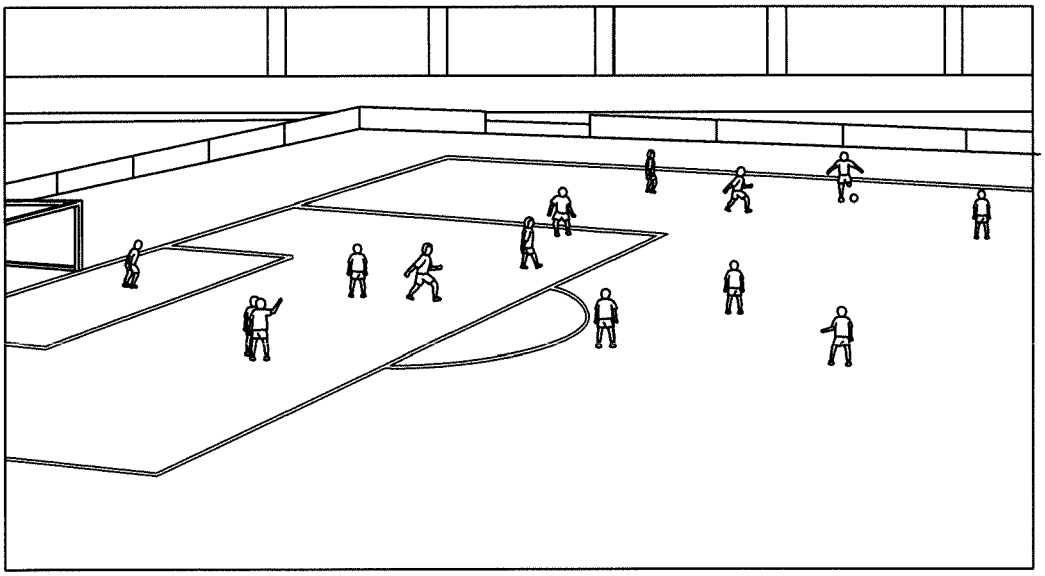


FIG.18

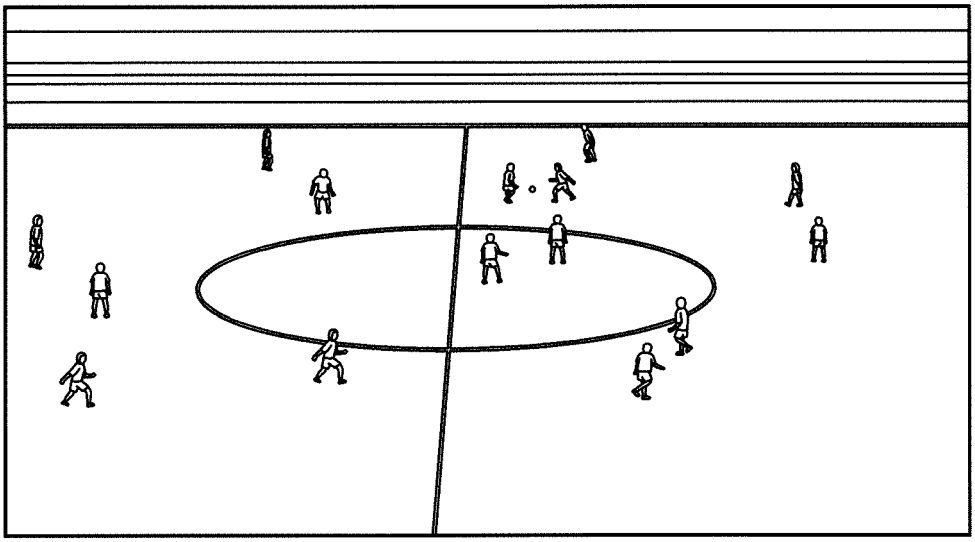


FIG.19

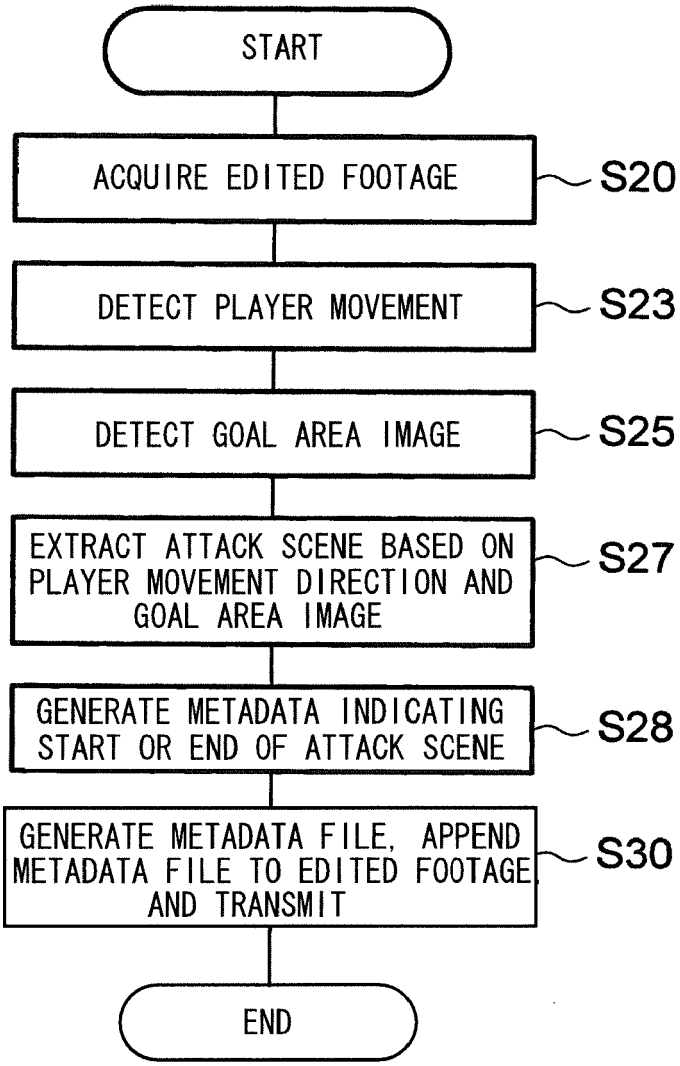
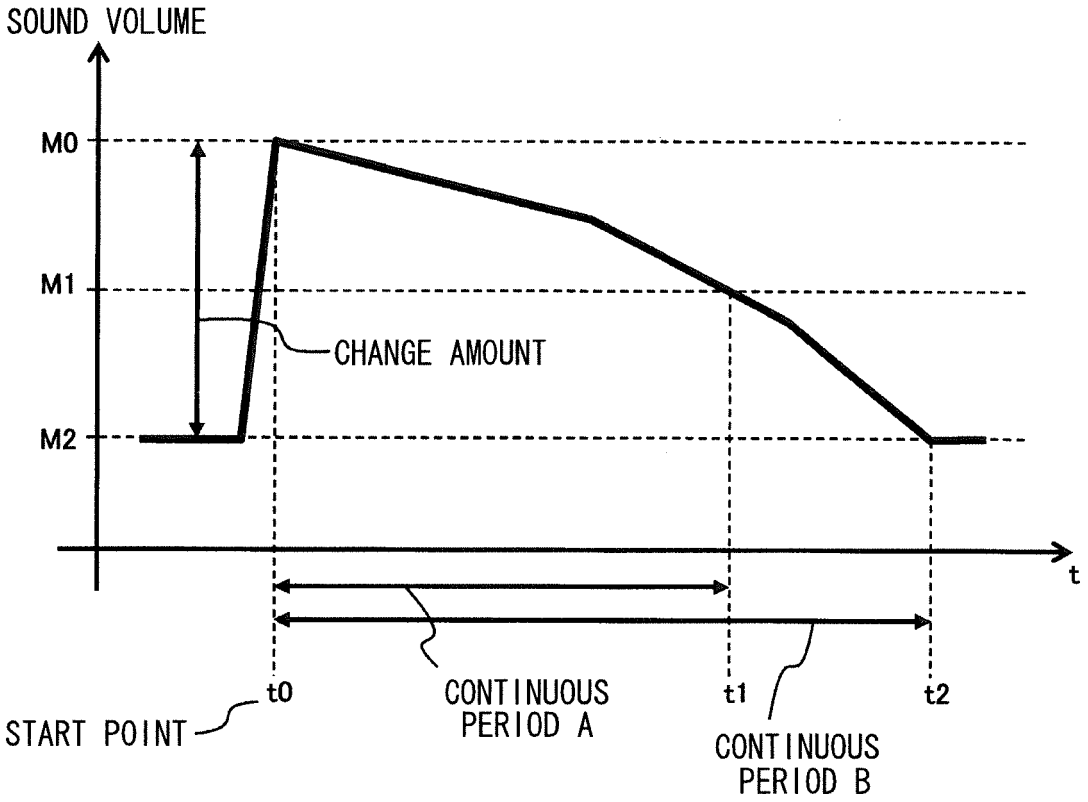

START
ACQUIRE EDITED FOOTAGE
S20
DETECT PLAYER MOVEMENT
S23
DETECT GOAL AREA IMAGE
S25
EXTRACT ATTACK SCENE BASED ON PLAYER MOVEMENT DIRECTION AND GOAL AREA IMAGE
S27
GENERATE METADATA INDICATING START OR END OF ATTACK SCENE
S28
GENERATE METADATA FILE, APPEND METADATA FILE TO EDITED FOOTAGE AND TRANSMIT
S30
END

FIG.20


SOUND VOLUME
M0
M1
M2
CHANGE AMOUNT
t
t0
t1
t2
START POINT
CONTINUOUS PERIOD A
CONTINUOUS PERIOD B

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 15 15 8182

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	YIHONG GONG ET AL: "Automatic parsing of TV soccer programs", PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON MULTIMEDIA COMPUTING AND SYSTEMS. WASHINGTON, MAY 15 - 18, 1995; [PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON MULTIMEDIA COMPUTING AND SYSTEMS], LOS ALAMITOS, IEEE COMP. SOC. PRESS, US, 15 May 1995 (1995-05-15), pages 167-174, XP010154594, DOI: 10.1109/MMCS.1995.484921 ISBN: 978-0-8186-7105-0 * page 167, left-hand column, paragraph 2 * * page 167, right-hand column, paragraph 3 * * page 168, right-hand column, paragraph 2 * * page 169, left-hand column, paragraph 1 * * page 171, right-hand column, paragraph 2-3 * * page 172, left-hand column, paragraph 1 * * section 5; page 172 *	1-15	INV. G11B27/034 G11B27/28 G06F17/30
	-----		TECHNICAL FIELDS SEARCHED (IPC)
X	EP 2 034 426 A1 (SONY CHINA LTD [CN]) 11 March 2009 (2009-03-11) * paragraph [0091]; figure 13 *	1-15	G11B G06F

A	WO 02/07164 A2 (UNIV COLUMBIA [US]; CHANG SHIH FU [US]; ZHONG DI [US]; KUMAR RAJ [US];) 24 January 2002 (2002-01-24) * pages 32,38 *	1-15	

	-/--		

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
The Hague	21 July 2015	Bruma, Cezar

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 15 15 8182

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	ASSFALG J ET AL: "Semantic annotation of soccer videos: automatic highlights identification", COMPUTER VISION AND IMAGE UNDERSTANDING, ACADEMIC PRESS, US, vol. 92, no. 2-3, 1 November 2003 (2003-11-01), pages 285-305, XP004472303, ISSN: 1077-3142, DOI: 10.1016/J.CVIU.2003.06.004 * the whole document *	1-15	
A	SHU-CHING CHEN ET AL: "Detection of Soccer Goal Shots Using Joint Multimedia Features and Classification Rules", INTERNET CITATION, 24 August 2003 (2003-08-24), XP002342134, Retrieved from the Internet: URL:Washington, DC, USA www.cs.fiu.edu/chens/PDF/mdm03.pdf [retrieved on 2005-08-25] * the whole document *	1-15	TECHNICAL FIELDS SEARCHED (IPC)
A,D	JP 2008 511186 A (MITSUBISHI DENKI KK, DIVAKARAN A, RADHAKRISHNAN R, XIONG Z) 10 April 2008 (2008-04-10) * abstract *	1-15	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
The Hague	21 July 2015	Bruma, Cezar

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 15 15 8182

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

21-07-2015

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 2034426 A1	11-03-2009	CN 101329766 A	24-12-2008
		EP 2034426 A1	11-03-2009
		JP 5420199 B2	19-02-2014
		JP 2008312215 A	25-12-2008
WO 0207164 A2	24-01-2002	AU 7596201 A	30-01-2002
		US 2004125877 A1	01-07-2004
		WO 0207164 A2	24-01-2002
JP 2008511186 A	10-04-2008	EP 1743265 A2	17-01-2007
		JP 2008511186 A	10-04-2008
		US 2006059120 A1	16-03-2006
		WO 2006022394 A2	02-03-2006

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- JP 2008511186 A **[0005]**